

## **Is language rhythm in the ear of the beholder? A sensorimotor synchronisation approach to the cross-linguistic study of rhythm**

Tamara Rathcke<sup>1</sup>, Chia-Yuan Lin<sup>1</sup>, Simone Falk<sup>2</sup>, Simone Dalla Bella<sup>2</sup>

<sup>1</sup>*University of Kent*, <sup>2</sup>*University of Montreal*

Rhythmic properties of speech and language have been controversially debated since 1980s (Dauer, 1983; Roach, 1982). The idea that linguistic rhythm can be reduced to two templates - “syllable-timing” and “stress-timing” (Abercrombie, 1967) – has been tested in numerous studies. However, empirical research to date has failed to provide evidence for an acoustic basis of the dichotomy (e.g., Arvaniti, 2009; Wiget et al., 2010), and the perceptual evidence in support of the idea is equally scarce (Arvaniti & Rodriguez, 2013; Miller, 1984). Following on from these criticisms, a wide range of proposals has been put forward, including ideas that language has a scale of rhythmicity (Kohler, 2009), or is even antirhythmic (Nolan & Jeon, 2014). Moreover, the quest for a rhythmic typology is sometimes viewed as Quixotic (Cummins, 2012:32). The main aim of the present study is to use a novel movement-based paradigm and provide evidence to what extent rhythm perception is bottom-up (i.e. guided by the acoustic signal) or top-down (i.e. shaped by the native prosodic system of the listener).

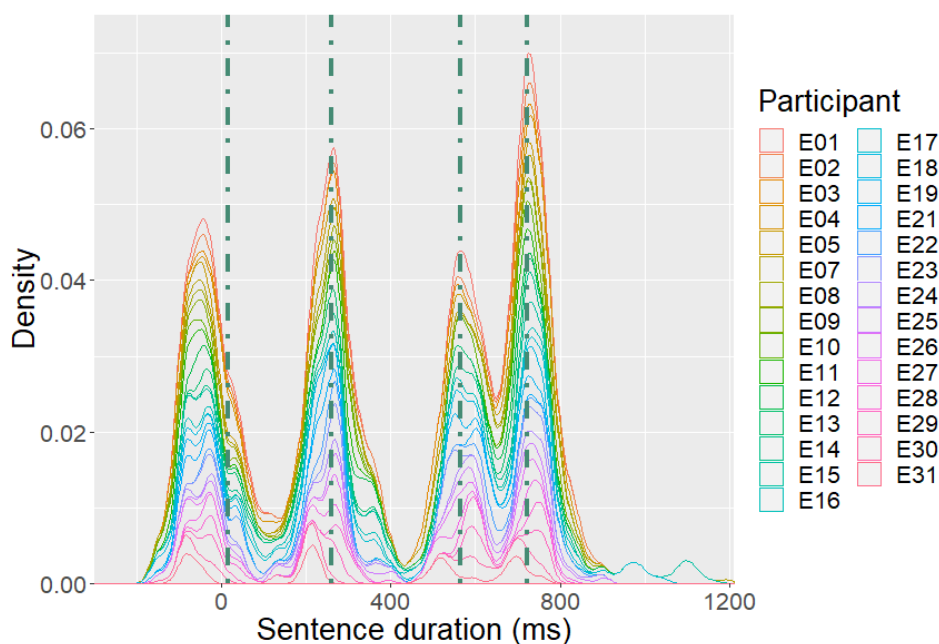
A sensorimotor synchronisation (SMS) paradigm was used in this research. SMS is a well-established laboratory tool to study beat tracking in non-verbal auditory and visual signals, by observing how a motor action is temporally coordinated with a pacing event (Aschersleben, 2002; Repp, 2005). French and English were chosen for the investigation as two prosodically distinct languages and the typical representatives of the two rhythmic templates, “syllable-timing” (French) and “stress-timing” (English). Thirty participants of each language were asked to tap in synchrony with the subjectively perceived beat of 20 sentences in their native and non-native language. The sentences varied in length and syntactic complexity, and were presented in a loop consisting of 20 repetitions. If rhythm perception arises bottom-up, the SMS data were expected to show evidence for some stable, acoustically defined events that would serve as synchronisation anchors in both groups of listeners. If rhythm perception is a top-down experience, the SMS data were expected to differ in a language-specific way.

The results showed some cross-linguistic similarities as both groups of listeners tended to time their taps relative to vowel onsets (see Fig. 1 for an example). However, there were also striking differences in the synchronisation patterns between the two groups of listeners, and were reflective of the listeners’ native prosodic system. English participants tended to tap with each vowel, though more frequently and accurately with strong (as opposed to weak) vowels. In contrast, French participants predominantly tapped with vowels in accented positions at the beginning and the end of an accentual phrase. We observed some rhythmic transfer in both listener groups when they tapped to their non-native language. Moreover, French listeners anticipated upcoming rhythmic events (especially pitch-accented vowels) while English listeners timed their taps more accurately to vowel onsets (especially in strong vowels).

These results, though currently limited to just two languages and read speech, first highlight that beat tracking in language is locked onto vowels. The cyclical production of vowel gestures in connected speech has been previously suggested as one of the main reasons why spoken language might be rhythmic in nature (Fowler & Tassinary, 1981). Our findings further indicate that linguistic rhythm may indeed be in the ear of the beholder, and the attempts to base a rhythmic typology entirely on acoustic properties of speech signals are likely to remain ill-advised (Cummins, 2012) whereas typological approaches that involve phonological features of prosodic systems (e.g. Jun, 2014) appear promising.

## References

- Abercrombie, D. (1967). *Elements of general phonetics*. (E. U. Press, Ed.). Edinburgh.
- Arvaniti, A. (2009). Rhythm, timing and the timing of rhythm. *Phonetica*, 66(1–2), 46–63.
- Arvaniti, A., & Rodriguez, T. (2013). The role of rhythm class, speaking rate, and F0 in language discrimination. *Laboratory Phonology*, 4(1), 7–38.
- Aschersleben, G. (2002). Temporal control of movements in sensorimotor synchronization. *Brain and Cognition*, 48(1), 66–79. <https://doi.org/10.1006/brcg.2001.1304>
- Cummins, F. (2012). Looking for rhythm in speech. *Empirical Musicology Review*, 7(1), 28–35.
- Dauer, R. M. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*, 11(1), 51–62.
- Fowler, C. A., & Tassinary, L. (1981). Natural measurement criteria for speech: The anisochrony illusion. In J. Long & A. Baddeley (Eds.), *Attention and performance, IX* (pp. 521–535). Hillsdale, NJ: Erlbaum.
- Jun, S.-A. (2014). Prosodic typology: by prominence type, word prosody, and macro-rhythm. In *Prosodic Typology II* (pp. 520–539).
- Kohler, K. J. (2009). Rhythm in speech and language: A new research paradigm. *Phonetica*, 66(1–2), 29–45. <https://doi.org/10.1159/000208929>
- Miller, M. (1984). On the perception of rhythm. *Journal of Phonetics*, 12, 75–83.
- Nolan, F., & Jeon, H. S. (2014). Speech rhythm: A metaphor? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1658), 20130396.
- Repp, B. H. (2005). Sensorimotor synchronization: A review of the tapping literature. *Psychonomic Bulletin and Review*, 12(6), 969–992.
- Roach, P. (1982). On the distinction between “stress-timed” and “syllable-timed” languages. In D. Crystal (Ed.), *Linguistics Controversies* (pp. 73–79). London: Edward Arnold.
- Wiget, L., White, L., Schuppler, B., Grenon, I., Rauch, O., & Mattys, S. L. (2010). How stable are acoustic metrics of contrastive speech rhythm? *The Journal of the Acoustical Society of America*, 127(3), 1559–1569.



**Figure 1.** Accumulated tapping frequencies of native English participants synchronising with the experimental sentence “I wove a yarn”. Dashed lines indicate vowel onsets.